Text Solutions pg. 189—197, #15, 17, 19, 27, 29, 35, 41

5. Cigarettes.

a) A linear model is probably appropriate. The residuals plot shows some initially low points, but there is not clear curvature.

b) 92.4% of the variability in nicotine level is explained by variability in tar content. (In other words, 92.4% of the variability in nicotine level is explained by the linear model.)

17. Another cigarette.

a) The correlation between tar and nicotine is $r = \sqrt{R^2} = \sqrt{0.924} = 0.961$. The positive value of the square root is used, since the relationship is believed to be positive. Evidence of the positive relationship is the positive coefficient of tar in the regression output.

b) The average nicotine content of cigarettes that are two standard deviations below the mean in tar content would be expected to be about 1.922 (2×0.961) standard deviations below the mean nicotine content.

c) Cigarettes that are one standard deviation above average in nicotine content are expected to be about 0.961 standard deviations (in other words, *r* standard deviations) above the mean tar content.

19. Last cigarette.

a) $\widehat{Nicotine} = 0.15403 + 0.065052(Tar)$ is the equation of the regression line that predicts nicotine content from tar content of cigarettes.

b)

$$\widehat{Nicotine} = 0.15403 + 0.065052(Tar)$$
$$\widehat{Nicotine} = 0.15403 + 0.065052(4)$$
$$\widehat{Nicotine} = 0.414$$

The model predicts that cigarette with 4 mg of tar will have about 0.414 mg of nicotine.

c) For each additional mg of tar, the model predicts an increase of 0.065 mg of nicotine.

d) The model predicts that a cigarette with no tar would have 0.154 mg of nicotine.

e)

$$\widehat{Nicotine} = 0.15403 + 0.065052(Tar)$$
$$\widehat{Nicotine} = 0.15403 + 0.065052(7)$$
$$\widehat{Nicotine} = 0.6094$$

The model predicts that a cigarette with 7 mg of tar will have 0.6094 mg of nicotine. If the residual is -0.5, the cigarette actually had 0.1094 mg of nicotine.

Text Solutions pg. 189—197, #15, 17, 19, 27, 29, 35, 41

27. SAT scores.

a) The association between SAT Math scores and SAT Verbal Scores was linear, moderate in strength, and positive. Students with high SAT Math scores typically had high SAT Verbal scores.

b) One student got a 500 Verbal and 800 Math. That set of scores doesn't seem to fit the pattern.

c) $r = 0.685$ indicates a moderate, positive association between SAT Math and SAT Verbal, but only because the scatterplot shows a linear relationship. Students who scored one standard deviation above the mean in SAT Math were expected to score 0.685 standard deviations above the mean in SAT Verbal. Additionally, $R^2 = (0.685)^2 = 0.469225$, so 46.9% of the variability in math score was explained by variability in verbal score.

d) The scatterplot of verbal and math scores shows a relationship that is straight enough, so a linear model is appropriate.

$$b_1 = \frac{r s_{Math}}{s_{Verbal}}$$
$$b_1 = \frac{(0.685)(96.1)}{99.5}$$
$$b_1 = 0.661593$$

$$\hat{y} = b_0 + b_1 x$$
$$\bar{y} = b_0 + b_1 \bar{x}$$
$$612.2 = b_0 + 0.661593(596.3)$$
$$b_0 = 217.692$$

The equation of the least squares regression line for predicting SAT Math score from SAT Verbal score is $\widehat{Math} = 217.692 + 0.662(Verbal)$.

e) For each additional point in verbal score, the model predicts an increase of 0.662 points in math score. A more meaningful interpretation might be scaled up. For each additional 10 points in verbal score, the model predicts an increase of 6.62 points in math score.

f)

$$\widehat{Math} = 217.692 + 0.662(Verbal)$$
$$\widehat{Math} = 217.692 + 0.662(500)$$
$$\widehat{Math} = 548.692$$

According to the model, a student with a verbal score of 500 was expected to have a math score of 548.692.

g)

$$\widehat{Math} = 217.692 + 0.662(Verbal)$$
$$\widehat{Math} = 217.692 + 0.662(800)$$
$$\widehat{Math} = 747.292$$

According to the model, a student with a verbal score of 800 was expected to have a math score of 747.292. She actually scored 800 on math, so her residual was 800 – 747.292 = 52.708 points

Text Solutions pg. 189—197, #15, 17, 19, 27, 29, 35, 41

29. SAT, take 2.

a) $r = 0.685$. The correlation between SAT Math and SAT Verbal is a unitless measure of the degree of linear association between the two variables. It doesn't depend on the order in which you are making predictions.

b) The scatterplot of verbal and math scores shows a relationship that is straight enough, so a linear model is appropriate.

$$b_1 = \frac{r s_{Verbal}}{s_{Math}}$$
$$b_1 = \frac{(0.685)(99.5)}{96.1}$$
$$b_1 = 0.709235$$

$$\hat{y} = b_0 + b_1 x$$
$$\bar{y} = b_0 + b_1 \bar{x}$$
$$596.3 = b_0 + 0.709235(612.2)$$
$$b_0 = 162.106$$

The equation of the least squares regression line for predicting SAT Verbal score from SAT Math score is: $\widehat{Verbal} = 162.106 + 0.709(Math)$

c) A positive residual means that the student's actual verbal score was higher than the score the model predicted for someone with the same math score.

d)

$$\widehat{Verbal} = 162.106 + 0.709(Math)$$
$$\widehat{Verbal} = 162.106 + 0.709(500)$$
$$\widehat{Verbal} = 516.606$$

According to the model, a person with a math score of 500 was expected to have a verbal score of 516.606 points.

e)

$$\widehat{Math} = 217.692 + 0.662(Verbal)$$
$$\widehat{Math} = 217.692 + 0.662(516.606)$$
$$\widehat{Math} = 559.685$$

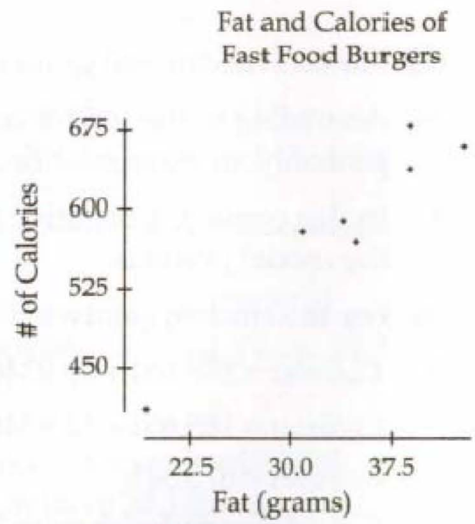
According to the model, a person with a verbal score of 516.606 was expected to have a math score of 559.685 points.

f) The prediction in part e) does not cycle back to 500 points because the regression equation used to predict math from verbal is a different equation than the regression equation used to predict verbal from math. One was generated by minimizing squared residuals in the verbal direction, the other was generated by minimizing squared residuals in the math direction. If a math score is one standard deviation above the mean, its predicted verbal score regresses toward the mean. The same is true for a verbal score used to predict a math score.

Text Solutions pg. 189—197, #15, 17, 19, 27, 29, 35, 41

35. Burgers.

a) The scatterplot of calories vs. fat content in fast food hamburgers is at the right. The relationship appears linear, so a linear model is appropriate.



Dependent variable is: **Calories**
No Selector
R squared = 92.3%   R squared (adjusted) = 90.7%
s = 27.33 with 7 - 2 = 5 degrees of freedom

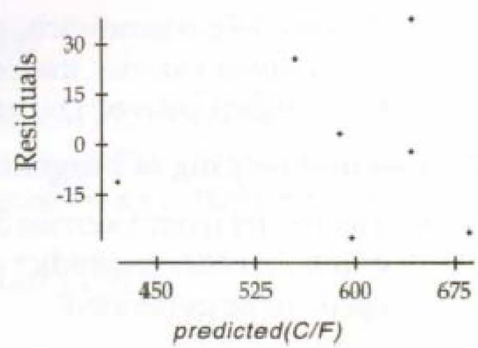
Source	Sum of Squares	df	Mean Square	F-ratio
Regression	44664.3	1	44664.3	59.8
Residual	3735.73	5	747.146	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	210.954	50.10	4.21	0.0084
Fat	11.0555	1.430	7.73	0.0006

b) From the computer regression output, $R^2 = 92.3\%$.
92.3% of the variability in the number of calories can be explained by the variability in the number of grams of fat in a fast food burger.

c) From the computer regression output, the regression equation that predicts the number of calories in a fast food burger from its fat content is: $\widehat{Calories} = 210.954 + 11.0555(Fat)$

d) The residuals plot at the right shows no pattern. The linear model appears to be appropriate.



e) The model predicts that a fat free burger would have 210.954 calories. Since there are no data values close to 0, this is an extrapolation outside the data and isn't of much use.

f) For each additional gram of fat in a burger, the model predicts an increase of 11.056 calories.

g) $\widehat{Calories} = 210.954 + 11.056(Fat) = 210.954 + 11.0555(28) = 520.508$
The model predicts a burger with 28 grams of fat will have 520.508 calories. If the residual is +33, the actual number of calories is 520.508 + 33 ≈ 553.5 calories.

41. El Niño.

a) The correlation between CO_2level and mean temperature is $r = R^2 = \sqrt{0.334} = 0.5779$.

b) 33.4% of the variability in mean temperature can be explained by variability in CO_2 level.

c) Since the scatterplot of CO_2 level and mean temperature shows a relationship that is straight enough, use of the linear model is appropriate. The linear regression model that predicts mean temperature from CO_2 level is: $Mean\hat{T}emp = 15.3066 + 0.004(CO_2)$

d) The model predicts that an increase in CO_2 level of 1 ppm is associated with an increase of 0.004 °C in mean temperature.

e) According to the model, the mean temperature is predicted to be 15.3066 °C when there is no CO_2 in the atmosphere. This is an extrapolation outside of the range of data, and isn't very meaningful in context, since there is always CO_2 in the atmosphere. We want to use this model to study the change in CO_2 level and how it relates to the change in temperature.

f) The residuals plot shows no apparent patterns. The linear model appears to be an appropriate one.

g)

$$Mean\hat{T}emp = 15.3066 + 0.004(CO_2)$$
$$Mean\hat{T}emp = 15.3066 + 0.004(364)$$
$$Mean\hat{T}emp = 16.7626$$

According to the model, the temperature is predicted to be 16.7626 °C when the CO_2 level is 364 ppm.

Text Solutions pg. 189—197, #15, 17, 19, 27, 29, 35, 41